AI Act
Tre questioni informatico-giuridico-filosofiche

Giovanni Sartor

AI Act: L'Intelligenza Artificiale al servizio della Pubblica Amministrazione - Impatto e Opportunità

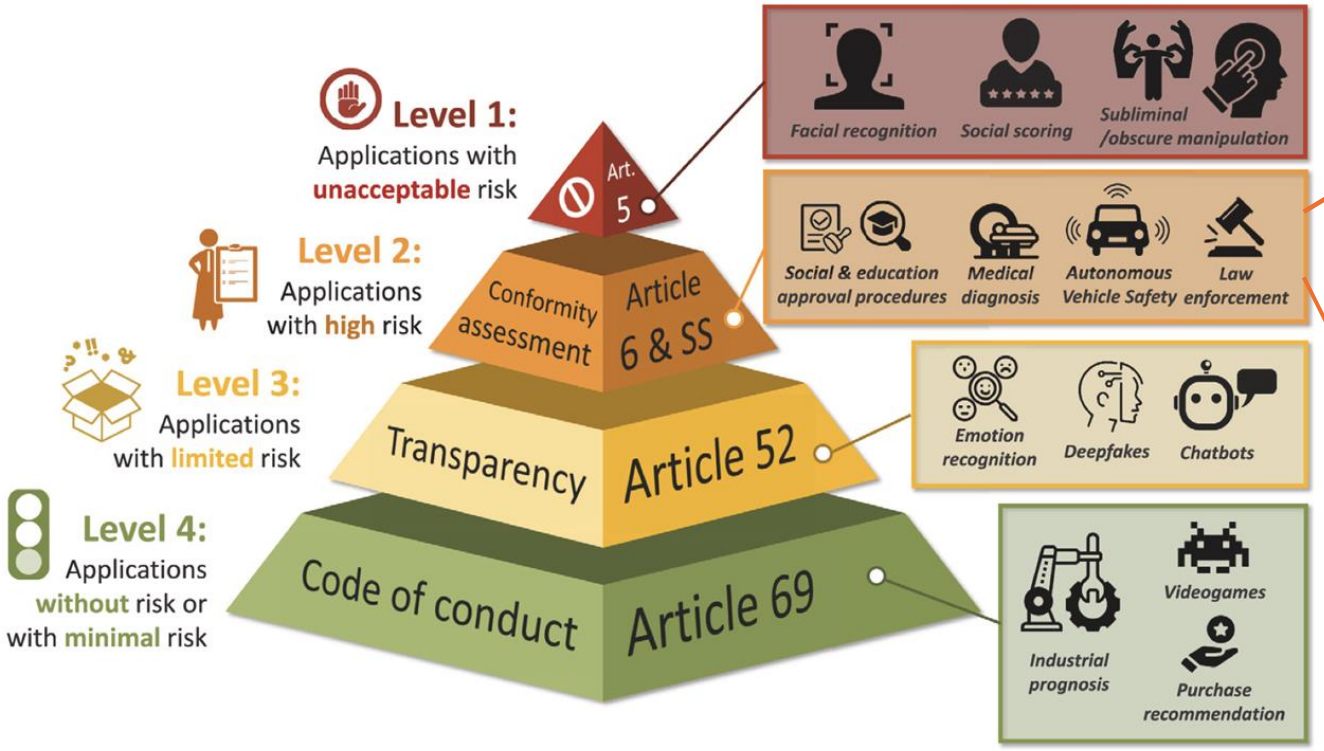
Bologna, 23 Aprile 2024

Questione 1. Quando può un sistema automatico dirsi artificialmente intelligente?

- Quando fa cose che richiederebbero intelligenza se fossero fatte da un uomo?
- Quanto usa certe tecnologie: sistemi basati sulla conoscenza, apprendimento automatico, statistica? (definizione nella proposta della Commissione)?
- Quando opera con «un variabile livello di autonomia» anziché con mera automaticità (Art 4 (1))?

- Problema “filosofico” che diventa giuridico: che cos’è l’autonomia? Un bancomat è autonomo? E un foglio Excel che fa qualche elaborazione statistica?
- Ma forse il concetto di IA non conta molto perché solo i sistemi «ad alto rischio» sono soggetti alle norme significative del regolamento

Un approccio basato sul rischio



Annex II legislation covers:

- Machinery
- Toys
- Lifts
- Radio Equipment
- Personal watercraft
- Medical devices
- Equipment used in potentially explosive atmosphere
- Pressure equipment
- Personally protective equipment
- Cableway installations
- Appliances burning gaseous fuels
- Civil aviation security
- 2- or 3-wheel vehicles and quadricycles
- Agricultural and forestry vehicles
- Marine equipment
- Interoperability of the rail system
- Motor vehicles and their trailers
- Civil aviation

Annex III categories of purposes:

- Biometric identification and categorisation of natural persons
- Management and operations of critical infrastructure
- Education and vocational training
- Employment, workers management and access to self-employment
- Access to and enjoyment of essential private services and public services and benefits
- Law enforcement
- Migration, asylum and border control management
- Administration of justice and democratic processes

Questione 2: Il rischio?

- L'AIA esprime una regolazione basata sul rischio. Ma cos'è il rischio?
- Rischio1 = evento dannoso
  - Es: Un rischio della guida autonoma è l'investimento di pedoni in zone urbane
- Rischio2 = La probabilità di un evento dannoso (Art. 4 (2))
  - Es: C'è un rischio elevato (Pr 1/1,000) che un'automobile autonoma usata in zone urbane investa un pedone nel corso di un anno di utilizzo
  - Ma se la probabilità di un evento dannoso non può essere determinate incertezza vs probabilità conosciuta (es. violazioni diritti fondamentali per AI)?
  - E se non si possono neppure individuare in modo esaustivo i possibili eventi dannosi (caso LLMs)?

Quando il rischio è alto?

- Rischio3 = costo atteso di un evento dannoso
  - Un possibile evento dannoso può dirsi un alto rischio quando il suo costo atteso è elevato (costo atteso = disutilità dell'evento * probabilità dell'evento)
    - Il rischio che l'automobile autonoma parcheggi al di là delle strisce che delimitano la zona consentita, anche se dotato di un elevate probabilità non può considerarsi un alto rischio in questo senso, perché il costo dell'evento è limitato (al massimo una multa, e l'utente se non è soddisfatto del parcheggio automatico può correggere a mano)
    - Il rischio di investire un pedone può invece considerarsi alto, quando la probabilità non sia estremamente bassa, data l'elevatissima disutilità dell'evento
- Un sistema può dirsi ad alto rischio quando il costo atteso di tutti i possibili eventi dannosi da esso generati (la somma del costo atteso di ciascuno di essi) sia elevata.
  - Problema: come determinarlo in modo ragionevole?
    - Presunzione relativa ex Art. 6 (1 e 2) , Eccezione ex. Art 6 (3).

Quando il rischio è accettabile?

- Gli elementi in gioco
  - Il costo atteso del sistema
  - I benefici del sistema

- Debbono esserci dei benefici
- I benefici debbono superare i costi

- Ma non basta
  - Bisogna che I costi siano necessari (per ottenere i benefici)

Accettabilità del rischio -> impossibilità di mitigazioni accettabili

- I rischi debbono essere limitati
  - usando ogni misura di mitigazione tale che
  - il costo della misura (diretto o in termini di riduzione dell'utilità del sistema) sia accettabile, cioè inferiore al beneficio (riduzione del rischio) che la misura comporta

- Non basta che l'automobile autonoma sia più sicura del guidatore umano
  - Bisogna che non sia possibile migliorarla così da ridurre gli incidenti (e.g. dotandola di , una telecamera migliore) con un costo accettabile (tale che il vantaggio nella riduzione del rischio superi il costo della misura)
- Problema:
  - Se i costi delle mitigazioni rendono il prodotto privo di mercato, esso non sarà messo realizzato e quindi i suoi benefici non saranno accessibili (Che fare se l'automobile estremamente sicura risulta così costosa o lenta che nessuno la vorrà e quindi nessuno la costruirà? Continueranno gli incidenti dovuti all'errore umano?)

Questione 3. L’equità (fairness)

- Tema giuridico: Non-discriminazione
  - Nessuno deve essere trattato sfavorevolmente sulla base di caratteristiche la cui considerazione è vietata(sesso, razza, opinioni politiche, ecc.)
  - O anche di caratteristiche correlate ma non essenziali (discriminazioni indiretta)

- Tema politico: Equa distribuzione dei vantaggi (e dei costi) dell’IA
  - Assicurare che tutti possano godere delle applicazioni di AI
  - Ridurre le diseguaglianza
    - o almeno considerare prioritariamente chi è sfavorito,
    - o almeno garantire a tutti un livello sufficiente di benessere

Equità e misure di parità statistica

- Approccio tecno-brutale
  - Equità (fairness) come parità statistica
    - *Parità demografica*: eguaglianza tra gruppi nella proporzione di classificazioni positive (o negative) rispetto alla popolazione
      - (es. in ogni gruppo si predice la malattia per stessa frazione X di pazienti; in ogni gruppo di candidati si prevede la capacità di svolgere il lavoro al livello richiesto per la stessa fraziiione Y di candidati))
    - «*Parità di opportunità*»: eguaglianza tra gruppi nella proporzione di classificazioni corrette positive (o negative) rispetto ap numero di individui positive (o negative).
      - (Es. in ogni gruppo i pazienti peri quali predice correttamente la malattia sono la stessa frazione X dei pazienti malati nel gruppo; in ogni gruppo, i candidatii quali si predice correttamente la capacità sono la stessa frazione Y di tutti i capaci nel gruppo)
    - «*Parità di trattamento*»: eguaglianza tra gruppi nel rapporto tra errori nelle classificazioni positive e nelle classificazioni negative
      - (ES ogni gruppo c'e' la stessa proporzione X tra pazienti sani di cui si è predetta la malattia e i pazienti malati di cui si è predetta la sanità; in ogni gruppo c'è la stessa proporzione Y tra gli incapaci cui si prevede la capacità, e i capaci di dui si prevede l'incapacità)
  - Scegliamo un criterio di parità statistica, o ottimizziamo in qualche modo la combinazione dei diversi criteri!
  - Bilanciamo le parità statistiche con accuratezza (riduciamo l'accuratezza delle predizioni nella misura in cui necessario per ottenere la parità statistica cui si mira).
- Ha senso procedere in questo modo? Quando?

Problemi dell’approccio tecno-brutale

- Problemi
  - Si trattano diversamente individui che sarebbero equivalenti rispetto alla predizione
    - (diversi individui con la stessa probabilità di avere un infarto sono classificati come sani o malati a seconda del gruppo cui appartengono, individui con la stessa probabilità di essere lavoratori capaci sono classificati come capaci o incapaci a seconda del gruppo)
  - Linguaggio fuorviante, che assimila complesse e dibattute questioni filosofiche (uguaglianza di opportunità o di trattamento) ad un semplice calcolo statistico
  - Possibilità di eguagliamento verso il basso
    - ottengo l’eguaglianza diminuendo le classificazioni vantaggiose per il gruppo che aveva ottenuto in media risultati migliori, senza avvantaggiare (in modo significativo) o addirittura danneggiando il gruppo che aveva ottenuto risultati peggiori
    - Ottengo un eguale ma più elevato numero previsioni errate per i pazienti malati in entrambi i gruppi
  - Non si discute la giustificazione etico-politica per l’eguagliamento che si ottiene nel contesto in esame.

Quale rimedio: Riflettere su …

- Che cosa vogliamo che il sistema predica?
  - Futura prestazione di lavoro?
  - Futuri risultati scolastici?
  - Futura propensione a restituire il credito concesso?
  - Futura (o presente) situazione di bisogno?
  - Futura (o presente) stato di malattia?
- Quali sono i fattori che consentono di ottenere le predizioni positive e negative più accurate?
- Quali obiettivi ci proponiamo mediante decisioni basate su questa previsione, secondo quali criteri. Nelle assunzioni:
  - Far sì che vengano assunte le persone che presumibilmente daranno prestazioni migliori?
  - Far sì che non vengano considerate caratteristiche ”proibite” né loro proxy (di regola coerente con il primo obiettivo)
  - Far sì che vengano superati gli svantaggi dovuti a condizioni famigliari e sociali? (es. abbassando la soglia richiesta per chi parte svantaggiato. Anche a scapito del primo obiettivo?)
  - Espandere la partecipazione a certe professioni di gruppi sottorappresentate?

Alcune osservazioni critiche…

- Se vogliamo adottare strategie di “azione affermativa”
  - dichiararlo espressamente, possibilmente stabilendo soglie diverse per valutazioni positive di candidati diversi gruppi, piuttosto che nascondere l’obiettivo egualitario all’interno di calcoli statistici
  - evitare risultatati matematicamente egualitari, ma assurdi nella pratica (eguagliamento verso il basso)
- Decisioni automatiche che consentano:
  - Chiarezza sui criteri adottati nelle scelte
  - Garanzia dell’attuazione dei criteri adottati
  - Aperto dibattito politico-sociale sugli obiettivi e i criteri

Conclusione

- Il regolamento sull'IA è un testo complesso (forse più del necessario) e di difficile attuazione
- E' necessario affrontare le questioni che solleva da una prospettiva interdisciplinare, che coinvolga, oltre a informatica e IA, discipline quali Economia, Diritto, Filosofia, Psicologia, ecc.
- L'attuazione del regolamento è una sfida difficile
- Si richiede la collaborazione di tutti!